The first approach was fairly simple, since there were no missing values there was no need of any imputation kind of thing. I tried removing all id columns which included geo id’s and building id. Next I used a random forest classifier with 200 estimators and used both criteria (entropy and gini). But the results were not that fine (I managed to achieve F1 score of 0.5918).

The second method was to add the geo\_id columns and used MLP classifier with different activation functions, solver, learning rate and tol levels and encoding the categorical columns I was able to achieve F1 scores of 0.6036.

The third method was similar to second one but this time I used Gradient Boosting Classifier with 500, 800, 1000 estimators respectively. Keeping the max depth to 2 and encoding the categorical columns I was able to achieve F1 scores of 0.5994, 0.6913 and 0.7066(best as of 16/06/2020).

With the score of 0.7066 I achieved the rank of 513(as on 16/06/2020).

The method is of course a best one as of now but in future I will be trying out different methods like XGBoost and others with some data analysis so as to achieve better score.